

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-153061

(43)Date of publication of application : 10.06.1997

(51)Int.Cl.

G06F 17/30

(21)Application number : 07-312669

(71)Applicant : FUJITSU LTD

(22)Date of filing : 30.11.1995

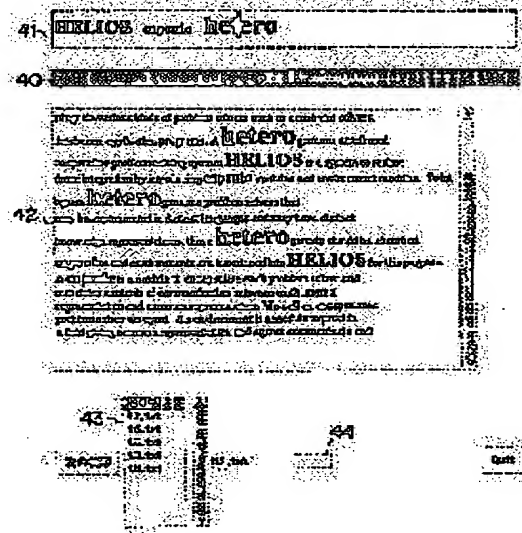
(72)Inventor : TSUDA HIROSHI

(54) METHOD AND DEVICE FOR DOCUMENT RETRIEVAL

(57)Abstract:

PROBLEM TO BE SOLVED: To enable a user to naturally input the degree of importance of a retrieval character string and to output the retrieval result in accordance with the inputted degree of importance.

SOLUTION: The retrieval character string is inputted to a window 41, and a mouse is clicked on the input character string to enlarge the font size, and the character string is emphasized in many stages. Next, a document including the input character string is retrieved, and the degree of relevance of this document is calculated in accordance with the degree of importance corresponding to the extent of emphasis of the input character string. File names of a certain number of documents are displayed in a window 43 in order of the degree of relevance, and contents of the document designated from them are displayed in a window 42. The retrieval character string is displayed in the window 42 in the same font size as input, and the user can recognize the relation between the retrieval character string and the document at sight.



LEGAL STATUS

[Date of request for examination] 28.11.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3606401

[Date of registration] 15.10.2004

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

BEST AVAILABLE COPY

(19) 日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11) 特許出願公開番号

特開平9-153061

(43) 公開日 平成9年(1997)6月10日

(51) Int.Cl.⁶

G 0 6 F 17/30

識別記号

庁内整理番号

F I

G 0 6 F 15/403

15/40

15/403

技術表示箇所

3 4 0 A

3 7 0 A

3 2 0 A

審査請求 未請求 請求項の数21 O L (全 20 頁)

(21) 出願番号 特願平7-312669

(22) 出願日 平成7年(1995)11月30日

(71) 出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番
1号

(72) 発明者 津田 宏

神奈川県川崎市中原区上小田中1015番地
富士通株式会社内

(74) 代理人 弁理士 大曾 義之 (外1名)

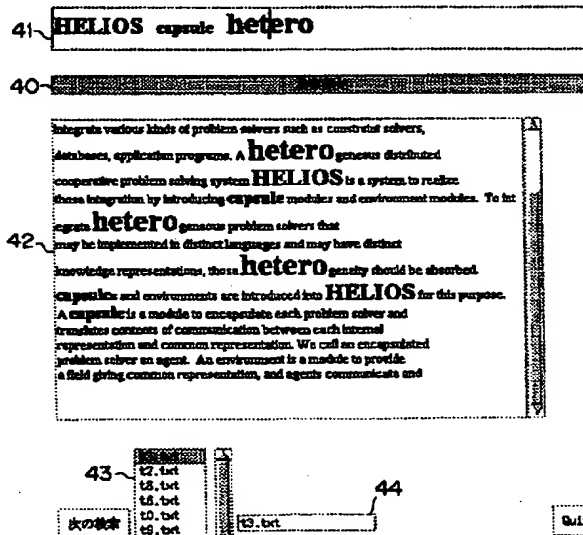
(54) 【発明の名称】 文書検索装置および方法

(57) 【要約】

【課題】 与えられた文字列を含む文書をデータベースから検索する情報処理装置において、ユーザが自然に検索文字列の重要度を入力することができ、入力された重要度に応じて検索結果を出力することを課題とする。

【解決手段】 ウィンドウ41内に検索文字列が入力され、入力文字列上でマウスをクリックすることによりフォントサイズが大きくなって、その文字列が多段階で強調される。次に、入力文字列を含む文書が検索され、入力文字列の強調度に応じた重要度からその文書の関連度が計算される。関連度の高いものから順に一定数の文書のファイル名がウィンドウ43に表示され、それらの中から指定された文書の内容がウィンドウ42に表示される。ウィンドウ42内では、検索文字列が入力時と同じフォントサイズで表示され、ユーザは検索文字列とその文書との関連を一目で認識することができる。

検索された文書の表示画面を示す図



1

【 特許請求の範囲】

【請求項1】 指定された文字列を検索キーとして、データベースから文書を検索する情報処理システムにおいて、検索文字列を視覚的に強調して入力する入力手段と、入力された前記検索文字列が含まれている文書を検索する検索手段と、検索された文書に含まれる前記検索文字列を、入力時に対応する強調度で強調して出力する出力手段とを備えることを特徴とする文書検索装置。

【請求項2】 前記出力手段は、入力時に指定された強調度で前記検索文字列を出力することを特徴とする請求項1記載の文書検索装置。

【請求項3】 前記入力手段は、入力された前記検索文字列を表示する表示手段と、該表示手段により表示された該検索文字列を指定して、該検索文字列の強調度を変更するポインティング手段とを備えることを特徴とする請求項1記載の文書検索装置。

【請求項4】 前記ポインティング手段は、前記検索文字列の大きさを前記強調度として指定する操作手段を備えることを特徴とする請求項2記載の文書検索装置。

【請求項5】 指定された文字列を検索キーとして、データベースから文書を検索する情報処理システムにおいて、検索文字列を視覚的に強調して入力する入力手段と、入力された前記検索文字列が含まれている文書を検索する検索手段と、検索された文書に含まれる前記検索文字列の強調度を重要度に対応させ、該重要度を用いて、該検索文字列と該検索された文書の間の関連を判定する関連度判定手段とを備えることを特徴とする文書検索装置。

【請求項6】 前記関連度判定手段は、前記検索文字列の強調度を前記重要度に変換し、該重要度を用いて前記検索された文書の関連度を計算し、該関連度の高い文書を優先的に検索結果とすることを特徴とする請求項5記載の文書検索装置。

【請求項7】 指定された文字列を検索キーとして、データベースから文書を検索する情報処理システムにおいて、検索文字列を入力する入力手段と、入力された前記検索文字列が含まれている文書を検索する検索手段と、検索結果を出力する出力手段と、出力された前記検索結果に含まれる情報を前記入力手段にフィードバックするフィードバック手段とを備えることを特徴とする文書検索装置。

【請求項8】 前記フィードバック手段は、出力された前記検索結果の任意の部分指定するポインティング手段を備え、該ポインティング手段により指定された文字列を新たな検索文字列として、前記入力手段にフィード

2

バックすることを特徴とする請求項7記載の文書検索装置。

【請求項9】 前記フィードバック手段は、出力された前記検索結果に含まれる前記検索文字列を指定して、該検索文字列の強調度を変更するポインティング手段を備え、該ポインティング手段により指定された強調度を前記入力手段にフィードバックすることを特徴とする請求項7記載の文書検索装置。

【請求項10】 前記ポインティング手段は、前記検索文字列の大きさを前記強調度として指定する操作手段を備えることを特徴とする請求項9記載の文書検索装置。

【請求項11】 前記入力手段は、前記フィードバック手段によりフィードバックされた情報に基づく新たな検索文字列を入力し、前記検索手段は該新たな検索文字列が含まれている文書を再び検索し、前記出力手段は再検索の結果を出力することを特徴とする請求項7記載の文書検索装置。

【請求項12】 検索された文書に含まれる前記検索文字列の重要度から該検索文字列と該検索された文書の間の関連を判定する関連度判定手段をさらに備え、前記出力手段は、該関連度判定手段の判定に基づいて前記検索結果を出力し、前記フィードバック手段は、該検索文字列の重要度を変更して前記入力手段にフィードバックし、該入力手段は変更された重要度を入力し、前記関連度判定手段は変更された重要度から前記関連を再度判定することを特徴とする請求項7記載の文書検索装置。

【請求項13】 情報処理システムにおいて、文字列を入力する入力手段と、あらかじめ用意された複数の強調度のうちの1つを用いて前記文字列を強調して、該文字列を含む文書を出力する出力手段とを備えることを特徴とする文書出力装置。

【請求項14】 指定された文字列を検索キーとして、データベースから文書を検索する情報処理に用いられる記憶媒体であって、検索文字列を視覚的に強調して入力する入力手段と、入力された前記検索文字列が含まれている文書を検索する検索手段と、検索された文書に含まれる前記検索文字列を、入力時に対応する強調度で強調して出力する出力手段とを備えることを特徴とする記憶媒体。

【請求項15】 指定された文字列を検索キーとして、データベースから文書を検索する情報処理に用いられる記憶媒体であって、検索文字列を視覚的に強調して入力する入力手段と、入力された前記検索文字列が含まれている文書を検索する検索手段と、検索された文書に含まれる前記検索文字列の強調度を重要度に対応させ、該重要度を用いて、該検索文字列と該検索された文書の間の関連を判定する関連度判定手段とを備えることを特徴とする記憶媒体。

50

3

【請求項16】 指定された文字列を検索キーとして、データベースから文書を検索する情報処理に用いられる記憶媒体であって、
 検索文字列を入力する入力手段と、
 入力された前記検索文字列が含まれている文書を検索する検索手段と、
 検索結果を出力する出力手段と、
 出力された前記検索結果に含まれる情報を前記入力手段にフィードバックするフィードバック手段とを備えることを特徴とする記憶媒体。

【請求項17】 文字列を入力する入力手段と、あらかじめ用意された複数の強調度のうちの1つを用いて前記文字列を強調して、該文字列を含む文書を出力する出力手段とを備えることを特徴とする記憶媒体。

【請求項18】 指定された文字列を検索キーとして、情報処理装置によりデータベースから文書を検索する方法において、
 検索文字列を視覚的に強調して入力し、
 入力された前記検索文字列が含まれている文書を検索し、
 検索された文書に含まれる前記検索文字列を、入力時に対応する強調度で強調して出力することを特徴とする文書検索方法。

【請求項19】 指定された文字列を検索キーとして、情報処理装置によりデータベースから文書を検索する方法において、
 検索文字列を視覚的に強調して入力し、
 入力された前記検索文字列が含まれている文書を検索し、

検索された文書に含まれる前記検索文字列の強調度を重要度に対応させ、
 該重要度を用いて、該検索文字列と該検索された文書の間の関連を判定することを特徴とする文書検索方法。

【請求項20】 指定された文字列を検索キーとして、情報処理装置によりデータベースから文書を検索する方法において、
 検索文字列を入力し、
 入力された前記検索文字列が含まれている文書を検索し、
 検索結果を出力し、
 出力された前記検索結果に含まれる情報を用いて再度文書を検索することを特徴とする文書検索方法。

【請求項21】 情報処理装置内の文書に含まれる文字列を、あらかじめ用意された複数の強調度のうちの1つを用いて強調して出力することを特徴とする文書出力方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、指定した文字列を含む文書をデータベースから検索する情報処理システム

4

において、検索キーとなる指定文字列の入力および検索結果の表示を行う文書検索装置およびその方法に関する。

【0002】

【従来の技術】計算機により機械的に可読な文書が増えている現代においては、与えられたキーワードとの関連度の高い文書を、通信ネットワークなどを通じて手軽に検索できるシステムが求められている。

【0003】従来の文書検索装置では、関連度の高い文書を検索するための仕組みとして、次の(1)、(2)、(3)、(4)のようなものが考案されている。

(1) 任意個のフリーターム(文字列)またはそれらの論理式を入力する検索システム

ユーザが任意個の検索文字列またはそれらの間の論理式を指定する。システムは、与えられた文字列が含まれている文書、または与えられた論理式が満たされている文書を検索する。そして、得られた文書の関連度をデータベース中の統計情報を利用して計算し、関連度の高い順に文書を表示する。WAI S(米国Wais社の検索システム)、TARGET(米国のDialog社の検索システム)などのシステムがある。

【0004】例えば、WAI Sにおいては、文書の関連度は次の4つの重み要素から計算されるとされている。ただし、その具体的な計算式は公開されていない。

(a) word weight: 検索語が文書中に出現した場合に加算されるweight。文書のheadlineに出現した場合が最もword weightが高く、すべての文字または最初の一文が大文字で出現した場合が次に高く、タイトル等を除く本文中にのみ出現した場合が最も低くなる。

【0005】(b) term weight: 文書中に頻繁に出てくる単語は、まれにしか出てこない単語に比べ、term weightが低い。

(c) proximity relationship: 複数の異なる検索語句が文書中で互いに近くに出現している場合に高いweightを与える。

【0006】(d) word density: 文書サイズに対する検索語の出現個数の比率

また、TARGETで用いられる関連度判定アルゴリズムでは、次のような要素が用いられ、関連度の高い方から50タイトルが検索結果として表示される。しかし、関連度の具体的な計算式は公開されていない。

【0007】(e) レコードに含まれる各検索語句の数
 (f) どの検索語句がレコードに現れているか

(g) 異なる検索語句同士がどれだけ近くに現れているか

(h) データベース中の検索語句の数

(2) 上記(1)のシステムに加えて、ユーザが検索語句の重要度を指定する検索システム

ユーザが任意個の検索文字列に、主として区間[0,

10

20

30

40

50

1]に属する実数による重要度を付与したり、またはそれらの重要度の論理式を指定したりする。このシステムでは上記(1)のシステムに加え、ユーザが指定した情報を用いて関連度計算を行うことが特徴である。いわゆる、重み付きブーリアン(weighted boolean)、または拡張ブーリアン(extended boolean)として、学会レベルで研究されている。

【0008】これらのうち最も簡単なものには、Fox and SharatによるMMM(Mixed Minand Max)モデルがある(Fox, E. A., and S. Sharat. 1986. "A Comparison of Two Methods for Soft Boolean Interpretation in Information Retrieval." Technical Report TR-86-1, Virginia Tech, Department of Computer Science.)。MMMモデルでは、文字列 w_i ($i=1, 2, \dots$)の重要度を d_i とすると、計算式 $w_1 \text{ AND } w_2$ の重要度 $= \min(d_1, d_2)$
 $w_1 \text{ OR } w_2$ の重要度 $= \max(d_1, d_2)$ により、重要度の論理式が定義される。文字列 w_1 、 w_2 を含む文書については、指定された計算式を用いて重要度が計算され、その文書がどの程度論理式を満たしているかが調べられる。そして、計算された値が大きいほど、関連度が大きいとみなされる。

(3)文章を入力し、それと類似したものを選ぶ検索システム

文字列ではなく、ある程度のサイズの文章を入力し、それに関連した文書をデータベースから検索する。入力文章の中からキーワードを適当に取り出し、似たようなキーワードの特徴を持つ文書を検索する。基本的な動作は、上記(1)や(2)のシステムとの違いはない。

(4)検索文字列を強調して出力するシステム

検索文字列を太字や下線で強調する、あるいは括弧で括弧などして、通常文字(地の文章)との区別をつけている。通常文字との区別がつけばいいだけなので、検索文字列の強調度は一段階のみである。このような出力処理に関する特許出願としては、「文字列強調出力方法」(特願昭61-080035、特開昭62-236746)「文書処理装置」(特願昭63-127103、特開平01-297768)「データ出力装置」(特願平02-328160、特開平04-195518)などがある。

【0009】

【発明が解決しようとする課題】しかしながら、上述のような従来の検索システムには次のような問題がある。(1)のシステムでは、ユーザが入力する検索語に重要度が付加されていないので、複数の検索語を入力した場合、それらはすべて同じレベルで処理される。このため、検索候補の文書の関連度の計算において用いられる情報が少なく、検索結果中にノイズが発生しやすい(適合率が低下する)。ここで、検索結果に発生するノイズとは、ユーザが検索を意図しなかった不要な文書を意味

する。例えば、入力された検索語のうち重要でないもののみが含まれる文書は、検索結果におけるノイズとなり得る。

【0010】(2)のシステムの場合、ユーザが重要度を0.2等の実数で指定するのは、入力に手間がかかるだけでなく、直観的にわかりにくいという問題がある。(3)のシステムでは、日本語の入力文章から単語を抽出する際に、英語と違ってどこで単語を切るかを決める解析処理が必要となる。これは、日本語の文章では、英語のように単語と単語の間にスペースが挿入されず、境界が明確ではないからである。単語を切る位置を誤ると、入力文章に関係のない文書が検索され、検索結果にノイズが発生する可能性がある。

【0011】また、検索語に重要度を加えて入力した場合、検索結果に含まれる検索語もその重要度に応じて表示することが望まれる。しかし、(4)のシステムのような文字列の一段階の強調出力では、重要度の異なる複数の文字列を区別できず、不十分である。

【0012】本発明は、与えられた文字列を含む文書をデータベースから検索するために、ユーザが自然に検索文字列の重要度を入力することができ、入力された重要度に応じて検索結果を出力する文書検索装置およびその方法を提供することを目的とする。

【0013】

【課題を解決するための手段】図1は、本発明の文書検索装置の原理図である。図1の文書検索装置は、指定された文字列を検索キーとして、データベース6から文書を検索する情報処理システム内に設けられ、入力手段1、検索手段2、関連度判定手段3、出力手段4、およびフィードバック手段5を備える。

【0014】入力手段1は、検索文字列を視覚的に強調して入力する。検索手段2は、入力された上記検索文字列が含まれている文書をデータベース6から検索する。

【0015】出力手段4は、検索された文書に含まれる上記検索文字列を、入力時に対応する強調度で強調して出力する。入力手段1は、検索文字列の強調度の大小を、例えばフォントサイズの大小などの文字列強調により視覚化する。強調度の視覚化は、例えばポインティング・デバイスを使ったGUI(グラフィカル・ユーザインタフェース)により、文字列のポイントとボタンのクリック操作で簡単に行うことができる。

【0016】出力手段4は、入力時の指定に対応する強調度により、検索された文書内の検索文字列を強調表示する。これにより、ユーザは検索結果の文書と検索文字列の間の関係を容易に把握することができる。

【0017】また、あらかじめ複数の強調度を用意しておけば、入力時に多段階の強調度のうちの1つを指定することができ、検索結果においても多段階の検索文字列の視覚化を行うことができる。

【0018】また、関連度判定手段3は、検索された文

7

書に含まれる上記検索文字列の強調度を重要度に対応させ、その重要度を用いて、上記検索文字列と上記検索された文書の間の関連を判定する。

【 0 0 1 9 】これにより、検索文字列の強調度がその重要度を表すことになり、入力時に指定された強調度が検索された文書の関連度に反映される。したがって、強調度を変更することにより、検索される文書の関連度を操作することが可能になる。

【 0 0 2 0 】また、フィードバック手段5 は、出力された上記検索結果に含まれる情報を入力手段1 にフィードバックする。出力手段4 が出力した検索結果において、例えば入力時と同様のGUI により、検索文字列およびその強調度を再指定することで、出力情報が入力手段1 にフィードバックされる。これにより、検索結果の文書内で新たな検索文字列を指定したり、既存の検索文字列の強調度を変更したりして、再検索を行うことが可能になる。

【 0 0 2 1 】例えば、図1 の入力手段1 は実施形態の図2 における入力部1 1 およびフィードバック指定部1 2 に対応し、検索手段2 は文字列検索部1 5 に対応し、関連度判定手段3 は関連度判定部1 4 に対応し、出力手段4 は出力部1 3 に対応し、フィードバック手段5 はフィードバック指定部1 2 に対応する。

【 0 0 2 2 】

【 発明の実施の形態】以下、図面を参照しながら本発明の実施の形態を詳細に説明する。図2 は、実施形態の文書検索装置の構成図である。図2 の文書検索装置は情報処理装置を用いて実現され、入力部1 1、フィードバック指定部1 2、出力部1 3、関連度判定部1 4、文字列検索部1 5、およびデータベース1 6 を備える。この文書検索装置における処理のフローを概説すると次のようになる。

【 0 0 2 3 】まず、ユーザは入力部1 1 に対して、キーボードおよびポインティング・デバイスにより、検索文字列とその強調度を入力する。入力された情報は文字列検索部1 5 に送られ、文字列検索部1 5 は、少なくとも一つの検索文字列の入っている文書をデータベース1 6 より選択する。

【 0 0 2 4 】次に、検索結果として得られた文書は関連度判定部1 4 に送られ、関連度判定部1 4 は、ユーザが入力した検索文字列の強調度に応じた重要度を用いて、文書の関連度を計算する。そして、算出された関連度の高い順に文書をソートして、ソート結果を出力部1 3 に送る。

【 0 0 2 5 】出力部1 3 は、ソート結果のうち関連度の高い順に上位一定数の文書をディスプレイ上に表示する。このとき、文書に含まれる検索文字列は、その強調度に応じて強調表示される。ユーザは、表示された文書の中の文字列を指定して強調度を変更するなどして、新たな検索文字列をフィードバック指定部1 2 に設定

8

し、文字列検索部1 5 に再検索を行わせることができる。

【 0 0 2 6 】図3 は、文書検索装置として用いられる情報処理装置の構成図である。図3 の情報処理装置は、データベース1 6、CPU (中央処理装置) 2 1、メモリ2 2、ポインティング・デバイス2 6、キーボード2 7、ディスプレイ2 8、およびそれらを結合するバス2 9 を備える。データベース1 6 は必ずしも情報処理装置の内部に設けられる必要はなく、外部のディスク装置上に設けてもよい。ディスク装置には、磁気ディスク、光ディスク、光磁気ディスクなどの任意のディスクを用いることができる。

【 0 0 2 7 】メモリ2 2 には、あらかじめ検索プログラム2 3 と関連度計算プログラム2 4 が格納され、これらはCPU 2 1 により実行される。CPU 2 1 は、検索プログラム2 3 を実行することにより入力部1 1、フィードバック指定部1 2、出力部1 3、および文字列検索部1 5 の機能を実現し、関連度計算プログラム2 4 を実行することにより関連度判定部1 4 の機能を実現する。また、一時的な検索結果2 5 をメモリ2 2 内に格納し、それを用いて処理を行う。

【 0 0 2 8 】キーボード2 7 は主として検索文字列の入力に用いられ、ポインティング・デバイス2 6 は主として文字列の強調度の指定に用いられる。文書検索装置の入力および出力に用いられる文字列の強調度として、例えば文字フォントの大きさを考える。フォントのサイズが大きいほど強調度が高いという設定は、日本語のみならず外国語においても通用する極めて自然なユーザインタフェースである。このような強調度を採用すれば、その実用化にあたっては、True Type フォント (ベクトルフォント) を扱うことのできるX - WindowやWindows などのOS (オペレーティングシステム) を搭載していれば容易である。ベクトルフォントとは、任意の大きさに拡大・縮小が可能なフォントを意味する。

【 0 0 2 9 】強調度としては、文字の色 (明るい色ほど強調度が高い) や書体などの他の属性を使用してもよい。しかし、これらの文字属性は、表示デバイスや使用言語の種類によって制限があり、汎用性にやや欠けている。そこで、本実施形態においては、フォントサイズの大小により強調度を指定することにする。

【 0 0 3 0 】強調度を指定する入力装置としては、ボタンが2 つ以上あるマウス等のポインティング・デバイス2 6 を使用する。図4 は、ポインティング・デバイス2 6 の一例であるマウスを示している。図4 のマウスは第1 ボタン3 1 と第2 ボタン3 2 を備える。ユーザは検索文字列 (検索語) をキーボード2 7 から入力した後、マウス3 0 を1 つの文字列の上に持っていく、第1 ボタン3 1 を押すことでその文字列のフォントサイズを大きくし、第2 ボタン3 2 を押すことでフォントサイズを小さくすることができる。

50

【0031】図5は、入力部11による入力処理のフローチャートである。図5において処理が開始されると、入力部11は、まずユーザがキーボード27またはマウス30を操作して検索開始を指示したかどうかを判定する(ステップS1)。検索開始の指示がなければ、次に、検索語がキーボード27から入力されたかどうかを判定し(ステップS2)、検索語が入力された場合はステップS1の処理に戻る。

【0032】図6は、ディスプレイ28上で入力された検索文字列の例を示している。図6において、検索文字列入力ウィンドウ41には、“HELIOS”、“capsule”、“hetero”の3つの検索語が入力されている。

【0033】次に、検索語が入力されずにマウス30が移動したことが分かった(ステップS3)、入力部11はマウス30の第1ボタン31がクリックされたかどうかを判定する(ステップS4)。ある検索語の上で第1ボタン31がクリックされた場合は、その単語のフォントサイズを拡大して(ステップS5)、ステップS1以降の処理を繰り返す。ステップS5においては、第1ボタン31のクリック回数、またはそれが押されている時間に応じて、適宜フォントサイズを大きくすればよい。

【0034】次に、第2ボタン32がクリックされたかどうかを判定し(ステップS6)、ある検索語の上で第2ボタン32がクリックされた場合は、その単語のフォントサイズを縮小して(ステップS7)、ステップS1以降の処理を繰り返す。ステップS7においても、第2ボタン32のクリック回数、またはそれが押されている時間に応じて、適宜フォントサイズを小さくすればよい。

【0035】また、第1ボタン31も第2ボタン32もクリックされなかった場合は、ステップS1の処理に戻る。そして、ステップS1で検索開始が指示されると、文書検索装置は検索処理を行う(ステップS8)。

【0036】図7は、このようにして強調度が指定された検索文字列を示している。図7の検索文字列入力ウィンドウ41において、3つの検索語はいずれもフォントサイズが拡大され、元のフォントに比べてより強調されていることが分かる。また、“HELIOS”と“hetero”は“capsule”よりもフォントサイズが大きく、強調度がより高いことを表している。

【0037】こうして、マウス30のクリック操作により検索文字列のフォントサイズが変化し、その強調度の大小がフォントサイズの大小により視覚化される。ユーザは、マウス30により文字列をポイントし、ボタンをクリックして、簡単に強調度の指定を行うことができる。図7において、ユーザが“検索開始”と記された操作ボタン40をクリックすると、これらの検索語をもとにして検索処理が開始される。

【0038】尚、文字列をポイントする方法として、マウス30の他にキーボード27のカーソルキーを用いて

もよく、また、他のポインティング・デバイスであるトラックボールや入力ペン等を用いてもよい。また、強調度の大小は、必ずしもフォントサイズの大小により視覚化する必要はなく、文字の色や書体の変化などの他の方法により視覚化してもよい。

【0039】図8は、文字列検索部15および関連度判定部14によるステップS8の検索処理のフローチャートである。図8において処理が開始されると、文字列検索部15は、まず指示された検索処理が前回と同じ検索文字列を含む再検索であって、検索文字列に変更がないかどうかを調べる(ステップS11)。そして、新たな検索や検索文字列の変更を伴う再検索の場合は、指定された検索文字列が1つでも含まれる文書をデータベース16から検索し、それらの文書の集合をDとする(ステップS12)。

【0040】次に、関連度判定部14は、検索文字列の強調度を用いて、集合D内の各文書の関連度を計算し(ステップS13)、関連度の大きい順にそれらをソートする(ステップS14)。そして、文書検索装置は、検索された文書に関する出力/フィードバック処理を行う(ステップS15)。

【0041】ステップS11において、指示された検索が再検索であり、検索文字列が不変で文字列の強調度の指定だけが前回と異なる場合には、文字列検索部15はデータベース16の検索を行わない。この場合、関連度判定部14は、メモリ22に保持されている前回の一時的検索結果25を利用して、ステップS13およびS14の処理を行う。ここで、一時的検索結果25は、前回検索された文書の集合Dを意味する。

【0042】ステップS13における関連度の計算では、入力時に指定された文字列の強調度がその文字列の重要度に変換され、その重要度を加味して文書の関連度が計算される。ここでは、強調度はフォントサイズにより指定されているので、フォントサイズに応じた重要度に基づいて関連度が計算される。

【0043】例えば、入力された n 個の検索文字列を s_1, s_2, \dots, s_n とし、それらの各フォントサイズを f_1, f_2, \dots, f_n とし、対象とする文書 j ($j=1, \dots, m$)に含まれる各文字列の個数を $c_{j1}, c_{j2}, \dots, c_{jn}$ とする。ここで、文書 j は、文字列 s_1, s_2, \dots, s_n の少なくとも1つを含むような m 個の文書のうちの1つである。

【0044】次に、フォントサイズ f_i ($i=1, \dots, n$)から重要度 $([0, 1]$ の実数)への関数を p とする。例えば、最小8ポイントから最大72ポイントまでの大きさのフォントが表示できるシステムの場合、関数 p としては、

II

$p(f_i) = ((f_i - 8) / 64)^2 \quad \dots (1)$
を用いることができる。

【0045】このとき、関連度 R_i を各文字列の重要度*

$$R_i = \sum_{j=1}^n p(f_{ij}) * c_{ij} \quad \dots (2)$$

(2)式は、重要度が大きい文字列が多く入っている文書ほど、関連度が高くなることを表している。

【0046】ステップS15において、出力部13は、入力時に指定された強調度により、文書中の検索文字列を強調してディスプレイ28に表示する。このときも、指定時と同様に、地の文章に対して数段階のフォントサイズによる強調表示が可能である。また、フィードバック指定部12は、入力部11と同様のインタフェースにより、出力結果の文書に対してフィードバック情報を加えることもできる。

【0047】図9は、出力部13およびフィードバック指定部12によるステップS15の出力/フィードバック処理のフローチャートである。また、図10、11、12、13、14は、出力/フィードバック処理における画面表示の例を示している。図9において処理が開始されると、出力部13は、まず検索結果の集合Dに含まれる文書のうち、関連度の大きなものから順に上位一定個数(例えば50個)の文書のファイル名を表示する(ステップS21)。次に、ユーザから検索終了の指示が入力されたかどうかを判定し(ステップS22)、終了指示がなければ、さらに再検索指示が入力されたかどうかを判定する(ステップS23)。

【0048】再検索の指示がなく、ユーザが画面上の1つのファイル名を選択した場合は(ステップS24)、選択されたファイル名に対応する文書の内容を表示する(ステップS25)。ユーザがファイル名をマウス30でクリックして読みたい文書を選択すれば、そのファイルの内容が表示される。このとき、入力時に指定した検索文字列は、指定時のフォントサイズで強調表示される。

【0049】図10は、図7の検索文字列を用いた検索により得られた文書の表示画面を示している。図10において、ウィンドウ43には関連度の大きな上位6個のファイル名が表示され、ウィンドウ44にはユーザが選択したファイル名t3.txtが表示されている。また、ウィンドウ42は、t3.txtの内容を示している。

【0050】ウィンドウ42内において、検索文字列“HELIOS”、“capsule”、“hetero”は、それぞれ検索文字列入力ウィンドウ41内の文字と同じフォントサイズで表示されており、ユーザはそれらの重要度を視覚的に容易に認識することができる。もし、文字の色や書体などの他の方法により強調度が指定されている場合は、検索結果においてもそれらに応じた表示を行う。

12

* $p(f_{ij})$ と個数 c_{ij} の積の和により定義すると、次式のようにになる。

【0051】次に、フィードバック指定部12は、表示された文書内で、マウス30のドラッグ操作により特定の文字列が選択されたかどうかを調べる(ステップS26)。文字列の選択がなければ出力部13によるステップS22の処理に戻り、文字列が選択されれば、次に、マウス30の第1ボタン31がクリックされたかどうかを判定する(ステップS27)。

【0052】図11は、図10の表示文書内で指定された文字列の例を示している。図11においては、マウスドラッグにより反転表示された“knowledge”が指定されており、これは最初の入力時の検索文字列とは異なることが分かる。

【0053】指定された文字列の上で第1ボタン31がクリックされた場合は、そのフォントサイズを拡大して(ステップS28)、ステップS27以降の処理を繰り返す。次に、第2ボタン32がクリックされたかどうかを判定し(ステップS29)、それがクリックされた場合はその文字列のフォントサイズを縮小して(ステップS30)、ステップS27以降の処理を繰り返す。ステップS28、S30におけるフォントサイズの変更方法は、それぞれ図5のステップS5、S7と同様である。

【0054】次に、マウスカーソルの移動があったかどうかを判定し(ステップS31)、移動がなければステップS27以降の処理を繰り返す。マウスカーソルが移動した場合は、ステップS26で指定された文字列を指定された強調度で検索文字列に加え、検索文字列入力ウィンドウ41に表示する。指定された文字列が既に検索文字列となっていれば、その強調度のみを変更する。こうして、検索文字列が更新され(ステップS32)、出力部13によるステップS22以降の処理が繰り返される。

【0055】このように、フィードバック指定部12によるステップS26からS32までの処理により、表示された文書中の文字列を新たな検索文字列としてデフォルトの強調度のままで追加したり、さらに検索文字列の強調度を変更したりすることができる。検索文字列の強調度を変更する場合は、ウィンドウ41内の該当する文字列をクリックしてもよく、ウィンドウ42内の文字列をクリックしてもよい。

【0056】図12は、図11で新たに指定された検索文字列“knowledge”の強調度を示している。図12においては、ウィンドウ41内に追加された文字列“knowledge”上で第1ボタン31がクリックされ、そのフォントサイズが拡大されている(ステップS28)。これ

に伴って、ウィンドウ42内の“knowledge”も拡大表示されている。

【0057】図13は、強調度が変更された既存の検索文字列の例を示している。図13において、ウィンドウ41内の最初に入力された検索文字列のうち、“HELIOS”と“hetero”上では第2ボタン32がクリックされ、そのフォントサイズが縮小されている（ステップS30）。また、文字列“capsule”上では第1ボタン31がクリックされ、そのフォントサイズが拡大されている（ステップS28）。これに伴って、ウィンドウ42内の“HELIOS”と“hetero”は縮小表示され、“capsule”は拡大表示されている。

【0058】こうして、ウィンドウ41内の検索文字列が更新され（ステップS32）、ステップS23で再検索が指示されると、出力部13は文字列検索部15に再検索を依頼する。これを受けて、文字列検索部15は、更新後の検索文字列をもとに再び図8の検索処理を行う（ステップS33）。

【0059】図14は、図13に示された検索文字列をもとに再検索を行った結果を表示した表示画面を示している。再検索は、操作ボタン40をクリックすることにより実行される。図14のウィンドウ43には関連度の大きな上位4個のファイル名が表示されている。これらのファイルの表示順序は、図10の場合に比べて変わっていることが分かる。また、ウィンドウ44にはユーザが選択したファイル名t3.txtが表示されており、ウィンドウ42にはt3.txtの内容が示されている。

【0060】ウィンドウ42内において、4つの検索文字列“HELIOS”、“capsule”、“hetero”、“knowledge”は、それぞれウィンドウ41内の文字と同じフォントサイズで表示されている。

【0061】このように、フィードバック指定部12により、出力部13が出力する検索結果が文字列検索部15にフィードバックされ、一旦検索された結果を用いて指定された検索文字列による再検索が行われる。これにより、ユーザは検索結果を効率良く利用しながら再検索を繰り返し、希望する文書を見つけることができる。

【0062】そして、ステップS22においてユーザから検索終了の指示があったとき、出力部13は処理を終了する。ここで、ユーザは図14の表示画面の“Quit”と記された操作ボタン46をクリックして、検索終了を指示することができる。また、“次の検索”と記された操作ボタン45をクリックすれば、表示内容を消去して新規の検索を開始することもできる。操作ボタン45がクリックされた場合は、入力部11は図5の入力処理を再び開始し、文書検索装置は次の検索を行う。

【0063】本発明の文書検索装置は、重み付きブーリアンや拡張ブーリアンを用いた検索システムのインタフェースとしても用いることができる。このようなシステ

ムにおいて、論理演算子による検索式（論理式）を入力した後、例えばマウス30で文字列を適宜クリックすることで、その文字列のフォントサイズを変更する。そして、（1）式のような変換式を用いて、フォントサイズを0から1までの重要度に変換すれば、論理式に基づく重要度を計算することが可能になる。フォントサイズ以外の他の強調度を用いた場合も同様である。

【0064】また、本発明の文書検索装置は、文書の速読支援ツールとしても使うことができる。強調度としてフォントサイズを利用した場合、重要度の高い文字列ほど大きなサイズで出力されるため、そのまわりの文章も読みやすくなる。強調表示されたそれらの文字列をたどることで、検索結果の文書を素速くながめることが可能になる。フォントサイズ以外の他の強調度を用いた場合も同様である。

【0065】

【発明の効果】本発明によれば、ユーザは自然なグラフィカル・ユーザインタフェースに基づいて文字列の強調度を指定し、入力文字列と関連度の高い文書を手軽に検索することができる。また、出力された検索結果は、詳しく読まずに眺めるだけでもある程度その重要性を認識することができる。

【0066】例えば、入力文字列の強調度をフォントサイズで表した場合、サイズが大きいものほど重要度が高いという設定は、直観的にも妥当なインタフェースになる。重要度を数値で指定する場合に比べて、画面上で文字列の変化を視覚的に捉えながら重要度を指定することができ、入力作業が容易である。

【0067】さらに、本発明は重み付きブーリアン検索のインタフェースや文書の速読支援ツールとしても用いることができる。

【図面の簡単な説明】

【図1】本発明の原理図である。

【図2】文書検索装置の構成図である。

【図3】情報処理装置の構成図である。

【図4】ポインティング・デバイスの例を示す図である。

【図5】入力処理のフローチャートである。

【図6】入力された文字列を示す図である。

【図7】強調度が指定された文字列を示す図である。

【図8】検索処理のフローチャートである。

【図9】出力／フィードバック処理のフローチャートである。

【図10】検索された文書の表示画面を示す図である。

【図11】文書内で新たに指定された文字列を示す図である。

【図12】新たに指定された文字列の強調度を示す図である。

【図13】強調度に変更された文字列を示す図である。

【図14】再検索結果の表示画面を示す図である。

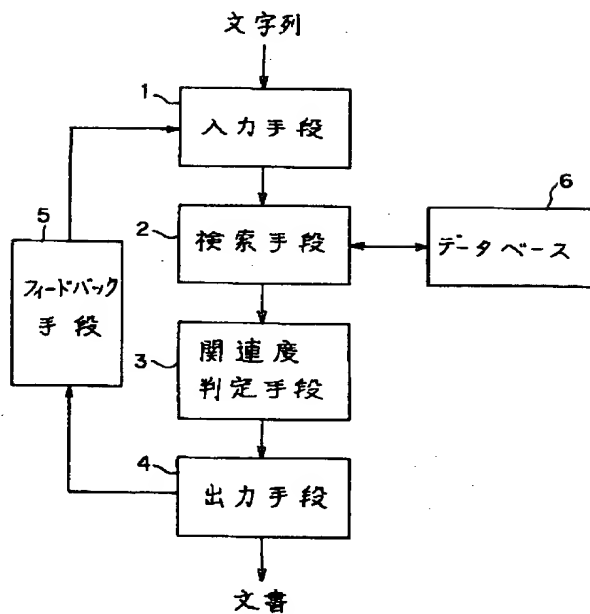
15

【 符号の説明】

- 1 入力手段
- 2 検索手段
- 3 関連度判定手段
- 4 出力手段
- 5 フィードバック手段
- 6、16 データベース
- 11 入力部
- 12 フィードバック指定部
- 13 出力部
- 14 関連度判定部
- 15 文字列検索部
- 21 CPU

【 図1】

本発明の原理図



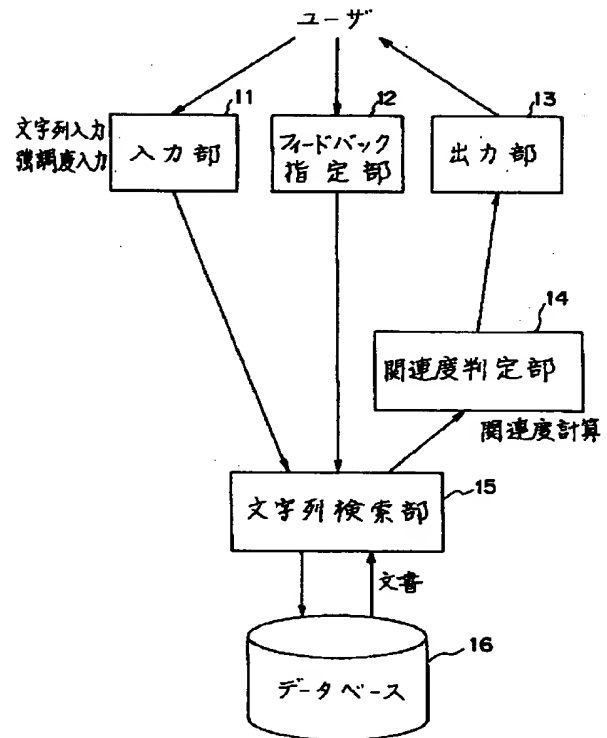
16

- 22 メモリ
- 23 検索プログラム
- 24 関連度計算プログラム
- 25 一時的検索結果
- 26 ポインティング・デバイス
- 27 キーボード
- 28 ディスプレイ
- 29 バス
- 30 マウス
- 31 第1 ボタン
- 32 第2 ボタン
- 40、45、46 操作ボタン
- 41、42、43、44 ウィンドウ

10

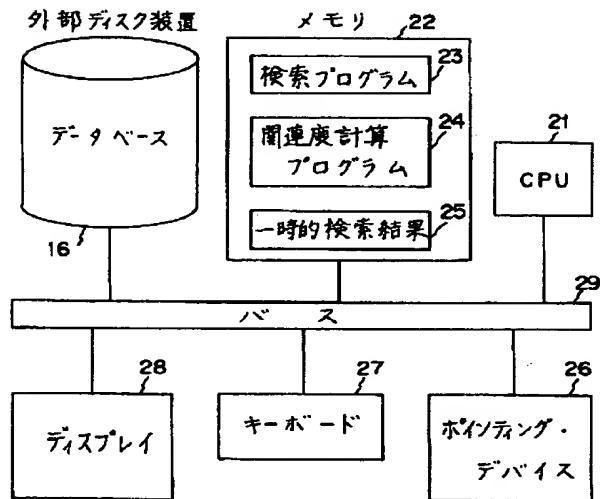
【 図2】

文書検索装置の構成図

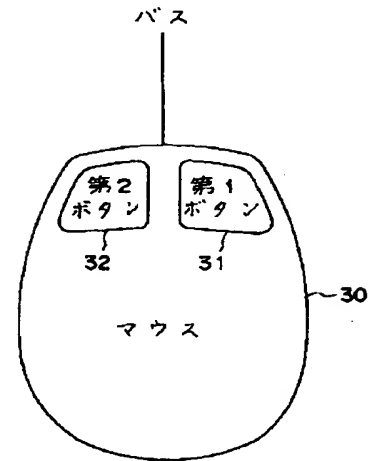


【 図3 】

情報処理装置の構成図

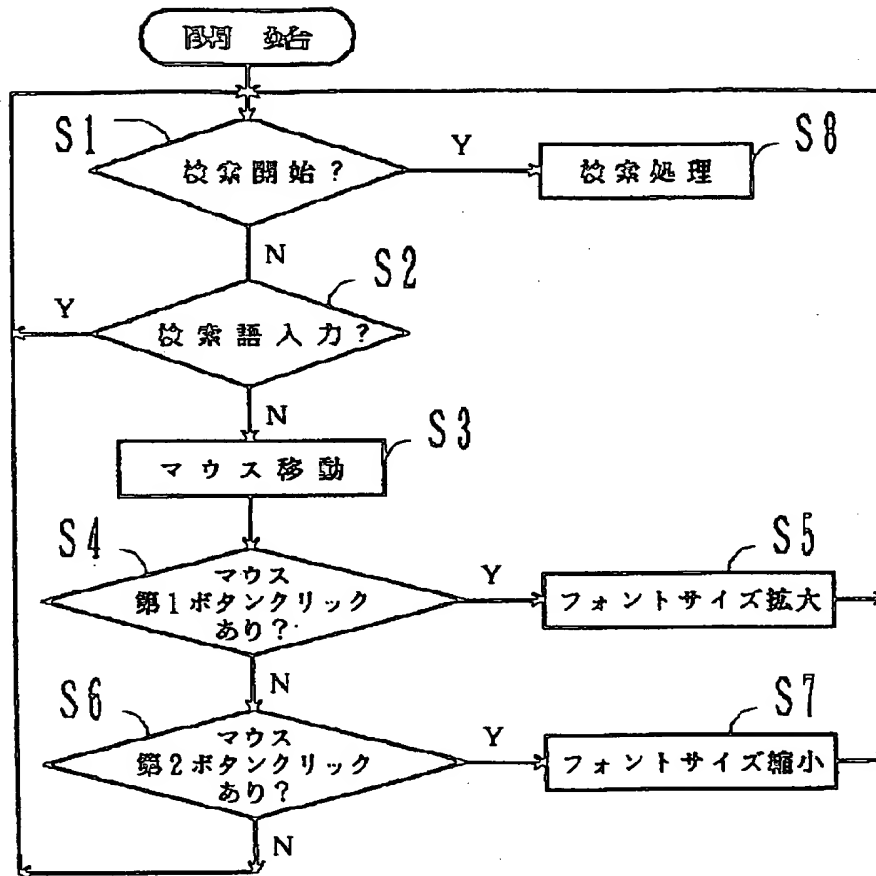


【 図4 】

ポインティング・デバイスの
例も示す図

【 図5 】

入 力 処 理 の フ ロ ー チ ャ ー ト



【 図6 】

入力された文字列を示す図

41 HELIOS capsule hetero

40 検索開始

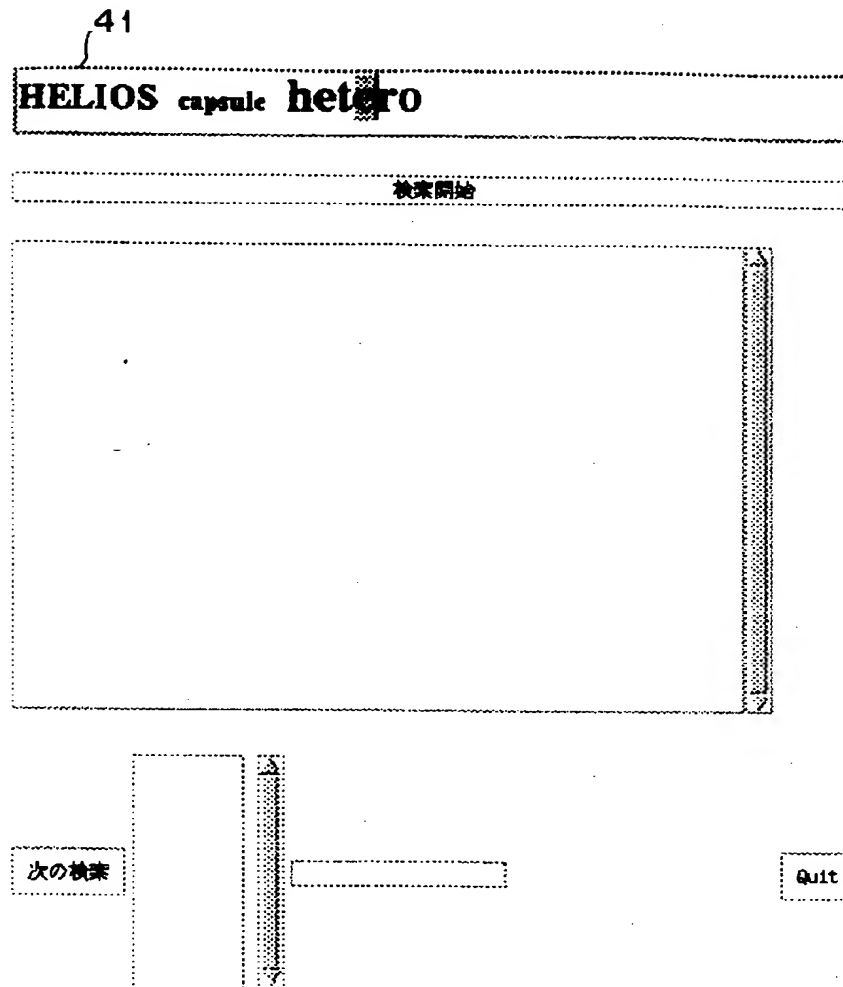
次の検索

Quit

The diagram illustrates a search interface. At the top, there is a text input field containing the string "HELIOS capsule hetero". Below this field is a button labeled "検索開始" (Start Search). To the left of the "検索開始" button is a label "41~". Below the input field and button is a large, empty rectangular area, likely for displaying search results. At the bottom of the interface, there are two buttons: "次の検索" (Next Search) on the left and "Quit" on the right. A vertical scrollbar is positioned between these two buttons.

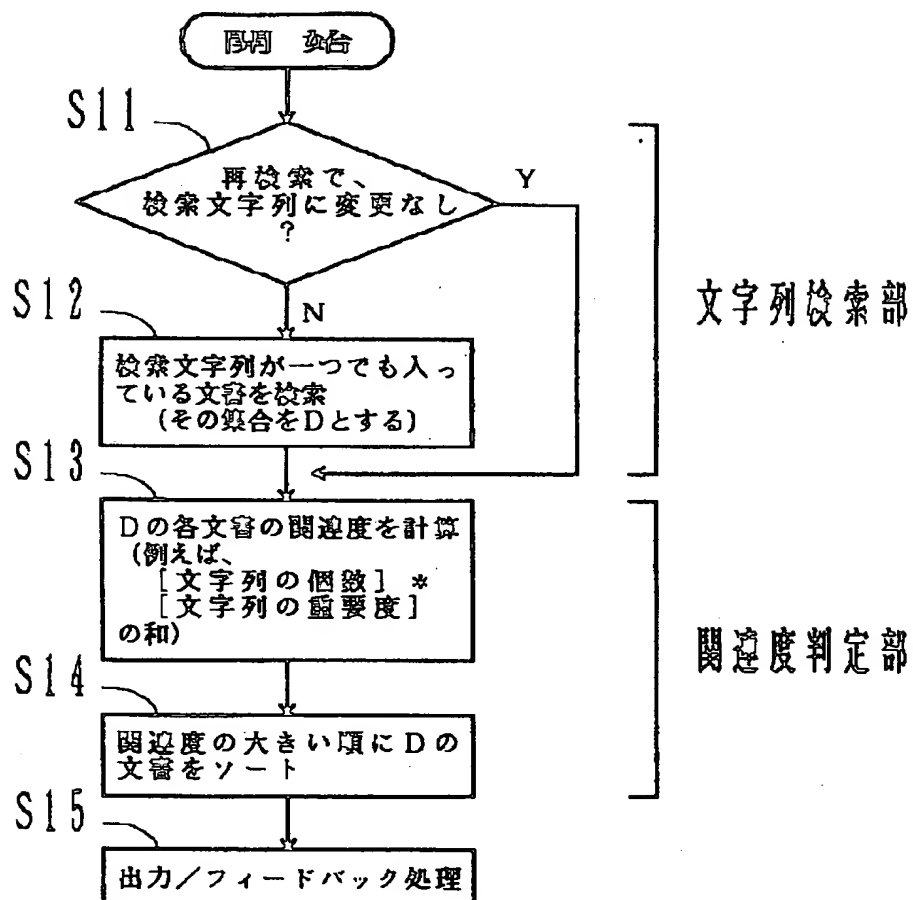
【 図7 】

強調度が指定された文字を示す図



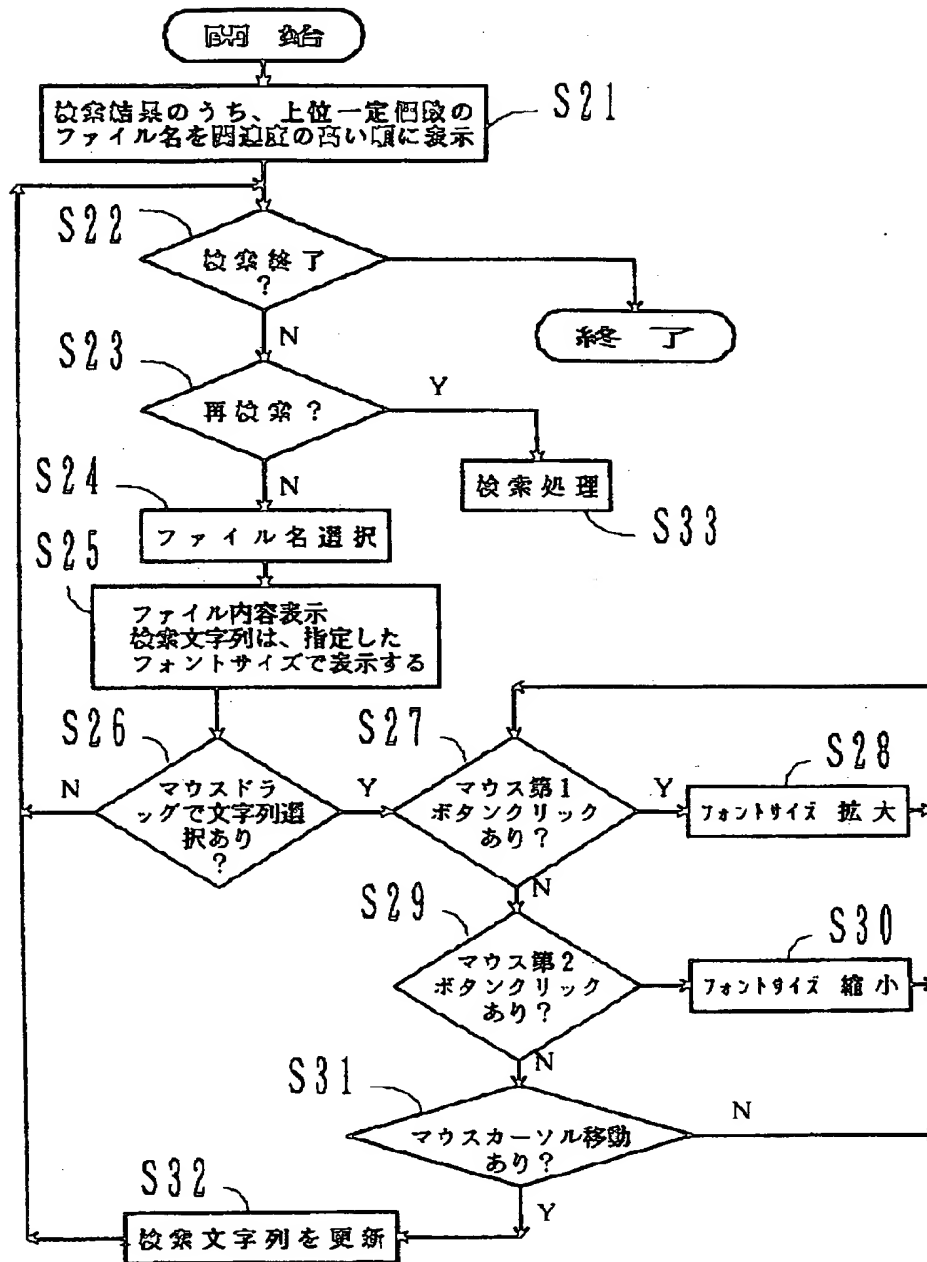
【 図8 】

検索処理のフローチャート



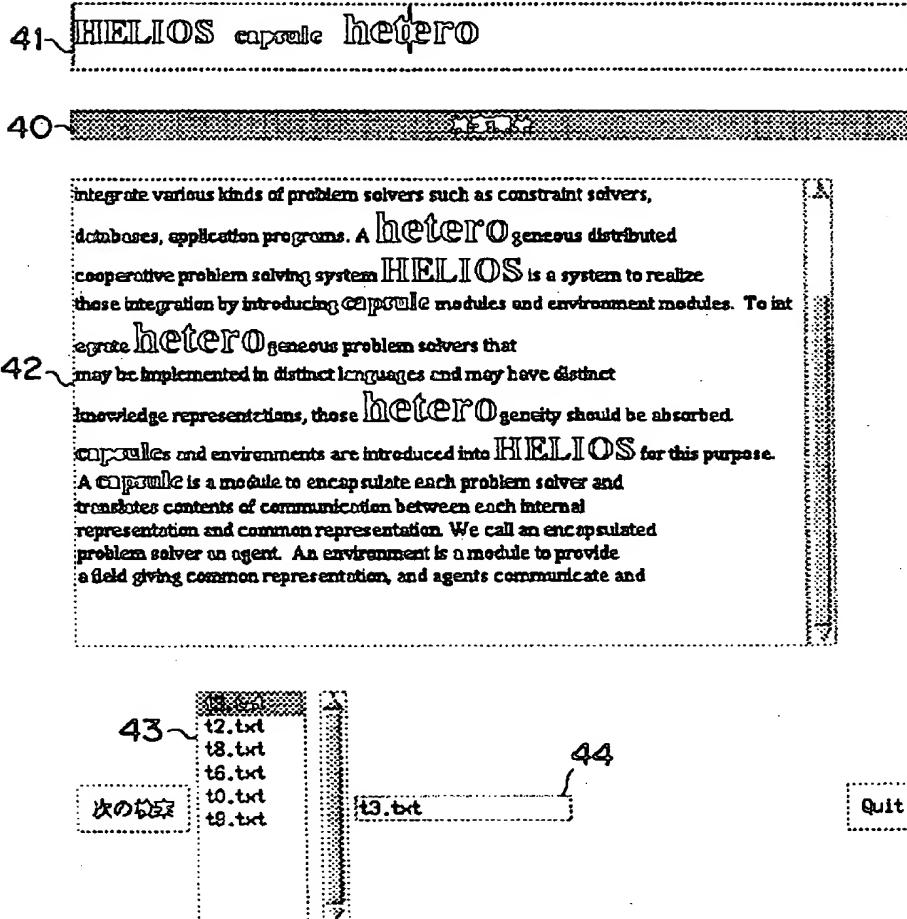
【 図9 】

出力 / フィードバック処理のフローチャート



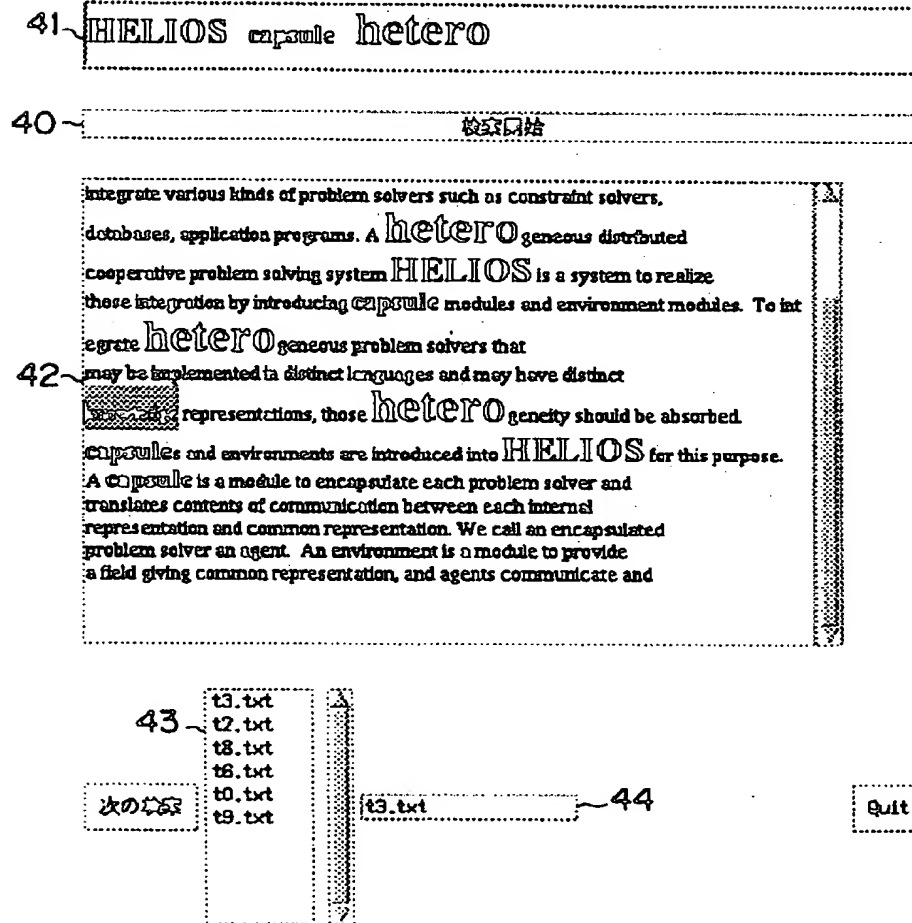
【 図10 】

検索された文書の表示画面を示す図



【 図 1 1 】

文書内で新たに指定された文字列を示す図



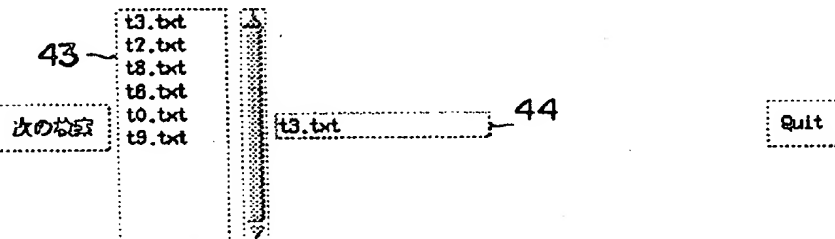
【 図1 2 】

新たに指定された文字列の強調度を示す図

41- HELIOS capsule hetero knowledge

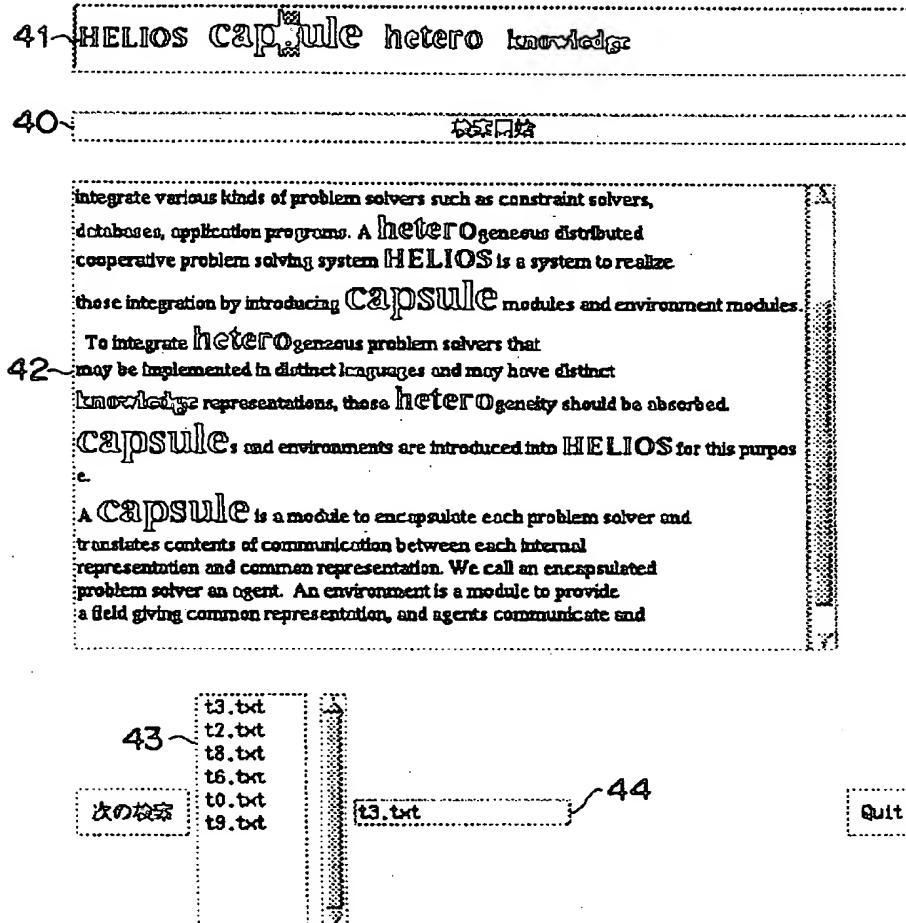
40- 検索開始

42- integrate various kinds of problem solvers such as constraint solvers, databases, application programs. A hetero geneous distributed cooperative problem solving system HELIOS is a system to realize those integration by introducing capsule modules and environment modules. To integrate hetero geneous problem solvers that may be implemented in distinct languages and may have distinct representations, those hetero geneity should be absorbed. capsules and environments are introduced into HELIOS for this purpose. A capsule is a module to encapsulate each problem solver and translates contents of communication between each internal representation and common representation. We call an encapsulated problem solver an agent. An environment is a module to provide a field giving common representation, and agents communicate and



【 図 1 3 】

強調度が変更された文字列を示す図



【 図1 4 】

再検索結果の表示画面を示す図

